TRhizo-urbanMicrobiome

Diversity & Composition Analyses

David Murray-Stoker

Contents

Load Packages

Load Data

```
## Load the root phyloseq object
root.phyloseq.reference <- read_rds(
  file = "data/root_decontam_phyloseq_reference.rds"
)

## Load the phyloseq object
soil.phyloseq.reference <- read_rds(
  file = "data/soil_decontam_phyloseq_reference.rds"
)

## Load the NJ tree
microbiome.NJ.tree <- read_rds(
  file = "microbiome_phylogenetic_tree/microbiome_NJ_tree.rds"
)

## Load the urbanization data
urbanization.data <- read_rds(
  file = "data/urbanization_data.rds"
)

## Load the carbon data
carbon.data <- read_csv(
  file = "data/urbanMicrobiome-carbon_data.csv",
  col_types = c("fnnn"),
  show_col_types = FALSE
)

## Load the nitrogen data
nitrogen.data <- read_csv(
  file = "data/urbanMicrobiome-nitrogen_data.csv",
  col_types = c("fnnn"),
  show_col_types = FALSE
)
```

Phyloseq Proccessing

```
## Rarefy to even sequencing depth
root.phyloseq.reference.rarefied <- rarefy_even_depth(
  root.phyloseq.reference,
  sample.size = 15000,
  rngseed = 9,
  trimOTUs = TRUE
)
# 2210 ASVs removed after being absent form all samples following rarefaction
# 12 samples removed as they had < 15000 reads

## Convert to tidyamplicons
root.tidyamplicon.reference.rarefied <- as_tidyamplicons(root.phyloseq.reference.rarefied)
```

```
## Rarefy to even sequencing depth
soil.phyloseq.reference.rarefied <- rarefy_even_depth(
  soil.phyloseq.reference,
  sample.size = 20000,
  rngseed = 9,
  trimOTUs = TRUE
)
# 1700 ASVs removed after being absent form all samples following rarefaction
# 4 samples removed as they had < 15000 reads

## Convert to tidyamplicons
soil.tidyamplicon.reference.rarefied <- as_tidyamplicons(soil.phyloseq.reference.rarefied)
```

```
## Merge root and soil phyloseq objects with the full microbiome NJ tree
microbiome.phyloseq.reference <- merge_phyloseq(
  root.phyloseq.reference.rarefied,
  soil.phyloseq.reference.rarefied,
  microbiome.NJ.tree
)

## Convert to tidyamplicons
microbiome.tidyamplicon.reference <- as_tidyamplicons(microbiome.phyloseq.reference)
```

Alpha Diversity

Quantify Alpha Diversity

```
## Extract the ASV abundance table
ASV.abundance.matrix <- abundances_matrix(microbiome.tidyamplicon.reference)

## Root ASV richness
ASV.richness <- specnumber(
  ASV.abundance.matrix
) %>%
  as.integer()

## Root Inverse Simpson's index
ASV.inverse.simpson <- diversity(
  ASV.abundance.matrix,
  index = "invsimpson"
) %>%
  as.numeric()

## Root ASV evenness
ASV.evenness <- ASV.inverse.simpson / ASV.richness %>%
  as.numeric()

## Root phylogenetic diversity
phylogenetic.diversity <- pd(
  ASV.abundance.matrix,
  microbiome.NJ.tree,
  include.root = FALSE
)

## Combine all alpha diversity estimates into one tibble
alpha.diversity.data <- tibble(
  Richness = ASV.richness,
  Inverse_Simpson = ASV.inverse.simpson,
  Evenness = ASV.evenness,
  Faiths_PD = phylogenetic.diversity$PD
)

## Combine root sample metadata and alpha diversity
alpha.diversity.cleaned.data <- samples(microbiome.tidyamplicon.reference) %>%
  full_join(urbanization.data, by = "Population") %>%
  bind_cols(alpha.diversity.data) %>%
  select(-c("sample", "sample_id")) %>%
  type_convert(
    col_types = c("fffininnnnninnn")
  )
```

ASV Richness GAMMs

```
## Richness by distance
ASV.richness.by.distance.GAMM <- gam(
  Richness ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Richness by HII
ASV.richness.by.HII.GAMM <- gam(
  Richness ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Richness by ISC
ASV.richness.by.ISC.GAMM <- gam(
  Richness ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)
```

Check Model Assumptions

```
## Root ASV richness-by-distance model diagnostics
check_model(ASV.richness.by.distance.GAMM)
# Visual check = assumptions met

## Root ASV richness-by-HII model diagnostics
check_model(ASV.richness.by.HII.GAMM)
# Visual check = assumptions met

## Root ASV richness-by-ISC model diagnostics
check_model(ASV.richness.by.ISC.GAMM)
# Visual check = assumptions met
```

ANOVAs

Table 1: ANOVA table for the ASV richness-by-distance GAMM. Adjusted R-squared = 0.676, deviance = 68.8. Compartment: F = 652.9, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.411	9	0.087	0.194
s(Distance):CompartmentSoil	2.939	9	2.848	0.000
s(Population)	7.687	34	0.292	0.123

Table 2: ANOVA table for the ASV richness-by-HII GAMM. Adjusted R-squared = 0.669, deviance = 68.1. Compartment: F = 637.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.505	9	0.138	0.156
s(Human_Influence_Index):CompartmentSoil	0.896	9	1.336	0.002
s(Population)	9.908	34	0.411	0.068

Table 3: ANOVA table for the ASV richness-by-ISC GAMM. Adjusted R-squared = 0.683, deviance = 69.0. Compartment: F = 665.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.551	9	0.135	0.138
s(Mean_ISC):CompartmentSoil	5.933	9	4.582	0.000
s(Population)	0.092	34	0.003	0.437

Inverse Simpson GAMMs

```
## Inverse Simpson by distance
inverse.simpson.by.distance.GAMM <- gam(
  Inverse_Simpson ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Inverse Simpson by HII
inverse.simpson.by.HII.GAMM <- gam(
  Inverse_Simpson ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Inverse Simpson by ISC
inverse.simpson.by.ISC.GAMM <- gam(
  Inverse_Simpson ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)
```

Check Model Assumptions

```
## Root inverse Simpson-by-distance model diagnostics
check_model(inverse.simpson.by.distance.GAMM)
# Visual check = assumptions met

## Root inverse Simpson-by-HII model diagnostics
check_model(inverse.simpson.by.HII.GAMM)
# Visual check = assumptions met

## Root inverse Simpson-by-ISC model diagnostics
check_model(inverse.simpson.by.ISC.GAMM)
# Visual check = assumptions met
```

ANOVAs

Table 4: ANOVA table for the inverse Simpson-by-distance GAMM. Adjusted R-squared = 0.820, deviance = 83.8. Compartment: F = 1331, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.002	9	0.000	0.845
s(Distance):CompartmentSoil	4.038	9	10.988	0.000
s(Population)	26.605	34	3.694	0.000

Table 5: ANOVA table for the inverse Simpson-by-HII GAMM. Adjusted R-squared = 0.822, deviance = 83.8. Compartment: F = 1347, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.001	9	0.000	0.989
s(Human_Influence_Index):CompartmentSoil	4.233	9	14.595	0.000
s(Population)	25.114	34	2.925	0.000

Table 6: ANOVA table for the inverse Simpson-by-ISC GAMM. Adjusted R-squared = 0.819, deviance = 83.6. Compartment: F = 1324, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.002	9	0.000	0.565
s(Mean_ISC):CompartmentSoil	4.372	9	10.431	0.000
s(Population)	26.064	34	3.523	0.000

ASV Evenness GAMMs

```
## Evenness by distance
ASV.evenness.by.distance.GAMM <- gam(
  Evenness ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Evenness by HII
ASV.evenness.by.HII.GAMM <- gam(
  Evenness ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Evenness by ISC
ASV.evenness.by.ISC.GAMM <- gam(
  Evenness ~ Compartment + s(Mean_ISC, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)
```

Check Model Assumptions

```
## Root ASV evenness-by-distance model diagnostics
check_model(ASV.evenness.by.distance.GAMM)
# Visual check = assumptions met

## Root ASV evenness-by-HII model diagnostics
check_model(ASV.evenness.by.HII.GAMM)
# Visual check = assumptions met

## Root ASV evenness-by-ISC model diagnostics
check_model(ASV.evenness.by.ISC.GAMM)
# Visual check = assumptions met
```

ANOVAs

Table 7: ANOVA table for the ASV evenness-by-distance GAMM. Adjusted R-squared = 0.632, deviance = 66.1. Compartment: F = 476.1, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.552
s(Distance):CompartmentSoil	0.000	9	0.000	0.377
s(Population)	25.038	34	2.787	0.000

Table 8: ANOVA table for the ASV evenness-by-HII GAMM. Adjusted R-squared = 0.632, deviance = 66.1. Compartment: F = 476.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.043	9	0.006	0.309
s(Human_Influence_Index):CompartmentSoil	0.001	9	0.000	0.443
s(Population)	25.031	34	2.786	0.000

Table 9: ANOVA table for the ASV evenness-by-ISC GAMM. Adjusted R-squared = 0.632, deviance = 66.1. Compartment: F = 476.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.859
s(Mean_ISC):CompartmentSoil	0.414	9	0.173	0.193
s(Population)	24.884	34	2.737	0.000

Faith's PD GAMMs

```
## Faith's PD by distance
faiths.PD.by.distance.GAMM <- gam(
  Faiths_PD ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Faith's PD by HII
faiths.PD.by.HII.GAMM <- gam(
  Faiths_PD ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)

## Faith's PD by ISC
faiths.PD.by.ISC.GAMM <- gam(
  Faiths_PD ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10)
    + s(Population, bs = "re", k = 35),
  data = alpha.diversity.cleaned.data,
  method = "REML"
)
```

Check Model Assumptions

```
## Faith's PD-by-distance model diagnostics
check_model(faiths.PD.by.distance.GAMM)
# Visual check = assumptions met

## Faith's PD-by-HII model diagnostics
check_model(faiths.PD.by.HII.GAMM)
# Visual check = assumptions met

## Faith's PD-by-ISC model diagnostics
check_model(faiths.PD.by.ISC.GAMM)
# Visual check = assumptions met
```

ANOVAs

Table 10: ANOVA table for the root Faith's PD-by-distance GAMM. Adjusted R-squared = 0.731, deviance = 74.3. Compartment: F = 839.1, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.533	9	0.151	0.147
s(Distance):CompartmentSoil	3.636	9	4.837	0.000
s(Population)	9.280	34	0.374	0.079

Table 11: ANOVA table for the root Faith's PD-by-HII GAMM. Adjusted R-squared = 0.722, deviance = 73.5. Compartment: F = 809.2, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.678	9	0.332	0.079
s(Human_Influence_Index):CompartmentSoil	0.927	9	2.258	0.000
s(Population)	12.636	34	0.593	0.022

Table 12: ANOVA table for the root Faith's PD-by-ISC GAMM. Adjusted R-squared = 0.737, deviance = 74.4. Compartment: F = 856.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.678	9	0.230	0.084
s(Mean_ISC):CompartmentSoil	6.443	9	6.677	0.000
s(Population)	1.149	34	0.035	0.382

Beta Diversity

Community Composition

Principal Coordinates Analyses

```
## Conduct PCoA for each distance matrix
# Bray-Curtis
BC.distance.PCoA <- ordinate(
  microbiome.phyloseq.reference.relativized,
  method = "MDS",
  distance = "bray"
)

# UniFrac
UniFrac.distance.PCoA <- ordinate(
  microbiome.phyloseq.reference.relativized,
  method = "MDS",
  distance = "unifrac"
)

# Weighted UniFrac
weighted.UniFrac.distance.PCoA <- ordinate(
  microbiome.phyloseq.reference.relativized,
  method = "MDS",
  distance = "wunifrac"
)


## Determine the percent of variation explained by axes 1-2
# Bray-Curtis = 38.5%
sum(BC.distance.PCoA$values$Eigenvalues[1:2]) / sum(BC.distance.PCoA$values$Eigenvalues)

# UniFrac = 24.7%
sum(UniFrac.distance.PCoA$values$Eigenvalues[1:2]) / sum(UniFrac.distance.PCoA$values$Eigenvalues)

# Weighted UniFrac = 55.7%
sum(weighted.UniFrac.distance.PCoA$values$Eigenvalues[1:2]) / sum(weighted.UniFrac.distance.PCoA$values
```

PERMANOVAs

```
## Set PERMANOVA metadata
sample.meta.data <- sample_data(microbiome.phyloseq.reference.relativized) %>%
  as_tibble()
```

```
## PERMANOVA by compartment
BC.distance.PERMANOVA <- adonis2(
  BC.distance ~ Compartment * Population,
  data = sample.meta.data,
  permutations = 1000
)
```

Table 13: Summary of the PERMANOVA comparing microbiome composition by compartment, population, and their interaction (Bray-Curtis distance).

Term	df	Sums-of-Squares	R2	F	P-value
Compartment	1	31.092	0.342	254.112	0.001
Population	34	16.638	0.183	3.999	0.001
Compartment:Population	34	11.156	0.123	2.682	0.001
Residual	262	32.057	0.352	NA	NA
Total	331	90.943	1.000	NA	NA

Bray-Curtis Distance

```r
## PERMANOVA by compartment
UniFrac.distance.PERMANOVA <- adonis2(
  UniFrac.distance ~ Compartment * Population,
  data = sample.meta.data,
  permutations = 1000
)
```

Table 14: Summary of the PERMANOVA comparing microbiome composition by compartment, population, and their interaction (UniFrac distance).

Term	df	Sums-of-Squares	R2	F	P-value
Compartment	1	19.358	0.215	116.340	0.001
Population	34	15.380	0.171	2.719	0.001
Compartment:Population	34	11.572	0.129	2.045	0.001
Residual	262	43.595	0.485	NA	NA
Total	331	89.904	1.000	NA	NA

UniFrac Distance

```r
## PERMANOVA by compartment
weighted.UniFrac.distance.PERMANOVA <- adonis2(
  weighted.UniFrac.distance ~ Compartment * Population,
  data = sample.meta.data,
  permutations = 1000
)
```

Table 15: Summary of the PERMANOVA comparing microbiome composition by compartment, population, and their interaction (weighted UniFrac distance).

Term	df	Sums-of-Squares	R2	F	P-value
Compartment	1	8.312	0.489	444.480	0.001
Population	34	2.299	0.135	3.616	0.001
Compartment:Population	34	1.495	0.088	2.352	0.001
Residual	262	4.900	0.288	NA	NA
Total	331	17.006	1.000	NA	NA

Weighted UniFrac Distance

Multivariate Dispersion

Dissimilarity Matrices

```
## Convert abundance to relative abundance
microbiome.phyloseq.reference.relativized <- transform_sample_counts(
  microbiome.phyloseq.reference,
  relative_abundance
)

## Conduct dissimilarity matrices
# Bray-Curtis
BC.distance <- distance(
  microbiome.phyloseq.reference.relativized,
  method = "bray"
)

# UniFrac
UniFrac.distance <- distance(
  microbiome.phyloseq.reference.relativized,
  method = "unifrac"
)

# Weighted UniFrac
weighted.UniFrac.distance <- distance(
  microbiome.phyloseq.reference.relativized,
  method = "wunifrac"
)
```

Quantify Multivariate Dispersion

```
## Calculate multivariate dispersion
# Bray-Curtis
BC.multivariate.dispersion <- betadisper(
  BC.distance,
  group = alpha.diversity.cleaned.data$Compartment,
  type = "centroid"
)$distances

# UniFrac
UniFrac.multivariate.dispersion <- betadisper(
  UniFrac.distance,
  group = alpha.diversity.cleaned.data$Compartment,
  type = "centroid"
)$distances

# Weighted UniFrac
weighted.UniFrac.multivariate.dispersion <- betadisper(
  weighted.UniFrac.distance,
  group = alpha.diversity.cleaned.data$Compartment,
  type = "centroid"
)$distances

## Compile the multivariate dispersion data
multivariate.dispersion.data <- tibble(
  Sequence_ID = alpha.diversity.cleaned.data$Sequence_ID,
  Population = alpha.diversity.cleaned.data$Population,
  Compartment = alpha.diversity.cleaned.data$Compartment,
  BC_Multivariate_Dispersion = BC.multivariate.dispersion,
  UniFrac_Multivariate_Dispersion = UniFrac.multivariate.dispersion,
  Weighted_UniFrac_Multivariate_Dispersion = weighted.UniFrac.multivariate.dispersion
) %>%
  group_by(Population, Compartment) %>%
  summarise(
    Mean_BC_Dispersion = mean(BC_Multivariate_Dispersion),
    Mean_UniFrac_Dispersion = mean(UniFrac_Multivariate_Dispersion),
    Mean_Weighted_UniFrac_Dispersion = mean(Weighted_UniFrac_Multivariate_Dispersion),
    .groups = "keep"
  ) %>%
  ungroup() %>%
  left_join(urbanization.data, by = "Population")
```

Bray-Curtis Multivariate Dispersion GAMs

```
##  Bray-Curtis multivariate dispersion-by-distance GAM
BC.multivariate.dispersion.by.distance.GAM <- gam(
  Mean_BC_Dispersion ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  Bray-Curtis multivariate dispersion-by-HII GAM
BC.multivariate.dispersion.by.HII.GAM <- gam(
  Mean_BC_Dispersion ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  Bray-Curtis multivariate dispersion-by-ISC GAM
BC.multivariate.dispersion.by.ISC.GAM <- gam(
  Mean_BC_Dispersion ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)
```

```
##  multivariate dispersion-by-distance model diagnostics
check_model(BC.multivariate.dispersion.by.distance.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-HII model diagnostics
check_model(BC.multivariate.dispersion.by.HII.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-ISC model diagnostics
check_model(BC.multivariate.dispersion.by.ISC.GAM)
# Visual check = assumptions met
```

Check Model Assumptions

Table 16: ANOVA table for the multivariate dispersion-by-distance GAM (Bray-Curtis distance). Adjusted R-squared = 0.630, deviance = 64.7. Compartment: F = 109.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	2.226	9	1.002	0.012
s(Distance):CompartmentSoil	0.000	9	0.000	0.866

Table 17: ANOVA table for the multivariate dispersion-by-HII GAM (Bray-Curtis distance). Adjusted R-squared = 0.582, deviance = 58.9. Compartment: F = 96.93, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.127	9	0.016	0.288
s(Human_Influence_Index):CompartmentSoil	0.001	9	0.000	0.369

Table 18: ANOVA table for the multivariate dispersion-by-ISC GAM (Bray-Curtis distance). Adjusted R-squared = 0.581, deviance = 58.7. Compartment: F = 96.72, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.001	9	0	0.367
s(Mean_ISC):CompartmentSoil	0.000	9	0	0.508

ANOVAs

UniFrac Multivariate Dispersion GAMs

```
##  UniFrac multivariate dispersion-by-distance GAM
UniFrac.multivariate.dispersion.by.distance.GAM <- gam(
  Mean_UniFrac_Dispersion ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  UniFrac multivariate dispersion-by-HII GAM
UniFrac.multivariate.dispersion.by.HII.GAM <- gam(
  Mean_UniFrac_Dispersion ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  UniFrac multivariate dispersion-by-ISC GAM
UniFrac.multivariate.dispersion.by.ISC.GAM <- gam(
  Mean_UniFrac_Dispersion ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)
```

```
##  multivariate dispersion-by-distance model diagnostics
check_model(UniFrac.multivariate.dispersion.by.distance.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-HII model diagnostics
check_model(UniFrac.multivariate.dispersion.by.HII.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-ISC model diagnostics
check_model(UniFrac.multivariate.dispersion.by.ISC.GAM)
# Visual check = assumptions met
```

Check Model Assumptions

Table 19: ANOVA table for the multivariate dispersion-by-distance GAM (UniFrac distance). Adjusted R-squared = 0.494, deviance 50.2. Compartment: F = 68.43, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	0.000	9	0.000	0.921
s(Distance):CompartmentSoil	0.012	9	0.001	0.319

Table 20: ANOVA table for the multivariate dispersion-by-HII GAM (UniFrac distance). Adjusted R-squared = 0.518, deviance = 53.0. Compartment: F = 71.72, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.063	9	0.007	0.306
s(Human_Influence_Index):CompartmentSoil	0.763	9	0.357	0.044

Table 21: ANOVA table for the multivariate dispersion-by-ISC GAM (UniFrac distance). Adjusted R-squared = 0.55, deviance = 57. Compartment: F = 76.87, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0.000	9	0.000	0.373
s(Mean_ISC):CompartmentSoil	2.124	9	0.933	0.015

ANOVAs

Weighted UniFrac Multivariate Dispersion GAMs

```
##  weighted UniFrac multivariate dispersion-by-distance GAM
weighted.UniFrac.multivariate.dispersion.by.distance.GAM <- gam(
  Mean_Weighted_UniFrac_Dispersion ~ Compartment
    + s(Distance, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  weighted UniFrac multivariate dispersion-by-HII GAM
weighted.UniFrac.multivariate.dispersion.by.HII.GAM <- gam(
  Mean_Weighted_UniFrac_Dispersion ~ Compartment
    + s(Human_Influence_Index, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)

##  weighted UniFrac multivariate dispersion-by-ISC GAM
weighted.UniFrac.multivariate.dispersion.by.ISC.GAM <- gam(
  Mean_Weighted_UniFrac_Dispersion ~ Compartment
    + s(Mean_ISC, by = Compartment, bs = "ts", k = 10),
  data = multivariate.dispersion.data,
  method = "REML"
)
```

```
##  multivariate dispersion-by-distance model diagnostics
check_model(weighted.UniFrac.multivariate.dispersion.by.distance.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-HII model diagnostics
check_model(weighted.UniFrac.multivariate.dispersion.by.HII.GAM)
# Visual check = assumptions met

##  multivariate dispersion-by-ISC model diagnostics
check_model(weighted.UniFrac.multivariate.dispersion.by.ISC.GAM)
# Visual check = assumptions met
```

Check Model Assumptions

Table 22: ANOVA table for the multivariate dispersion-by-distance GAM (weighted UniFrac distance). Adjusted R-squared = 0.656, deviance = 67.2. Compartment: F = 122.4, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Distance):CompartmentRoot	2.255	9	1.122	0.008
s(Distance):CompartmentSoil	0.000	9	0.000	0.987

Table 23: ANOVA table for the multivariate dispersion-by-HII GAM (weighted UniFrac distance). Adjusted R-squared = 0.605, deviance = 61.0. Compartment: F = 106.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Human_Influence_Index):CompartmentRoot	0.001	9	0	0.355
s(Human_Influence_Index):CompartmentSoil	0.000	9	0	0.614

Table 24: ANOVA table for the multivariate dispersion-by-ISC GAM (weighted UniFrac distance). Adjusted R-squared = 0.605, deviance = 61.0. Compartment: F = 106.6, P < 0.001.

Term	EDF	Ref. df	F	P-value
s(Mean_ISC):CompartmentRoot	0	9	0	0.471
s(Mean_ISC):CompartmentSoil	0	9	0	0.582

ANOVAs

Export Data

```
## Set root PCoA data
root.PCoA.vectors <- tibble(
  Root_BC_PCoA_1 = BC.distance.PCoA$vectors[, 1],
  Root_BC_PCoA_2 = BC.distance.PCoA$vectors[, 2],
  Root_UniFrac_PCoA_1 = UniFrac.distance.PCoA$vectors[, 1],
  Root_UniFrac_PCoA_2 = UniFrac.distance.PCoA$vectors[, 2],
  Root_Weighted_UniFrac_PCoA_1 = weighted.UniFrac.distance.PCoA$vectors[, 1],
  Root_Weighted_UniFrac_PCoA_2 = weighted.UniFrac.distance.PCoA$vectors[, 2]
) %>%
  bind_cols(sample.meta.data) %>%
  select(Population:Compartment, Root_BC_PCoA_1:Root_Weighted_UniFrac_PCoA_2) %>%
  filter(Compartment == "Root") %>%
  select(-Compartment)

## Set soil PCoA data
soil.PCoA.vectors <- tibble(
  Soil_BC_PCoA_1 = BC.distance.PCoA$vectors[, 1],
  Soil_BC_PCoA_2 = BC.distance.PCoA$vectors[, 2],
  Soil_UniFrac_PCoA_1 = UniFrac.distance.PCoA$vectors[, 1],
  Soil_UniFrac_PCoA_2 = UniFrac.distance.PCoA$vectors[, 2],
  Soil_Weighted_UniFrac_PCoA_1 = weighted.UniFrac.distance.PCoA$vectors[, 1],
  Soil_Weighted_UniFrac_PCoA_2 = weighted.UniFrac.distance.PCoA$vectors[, 2]
) %>%
  bind_cols(sample.meta.data) %>%
  select(Population:Compartment, Soil_BC_PCoA_1:Soil_Weighted_UniFrac_PCoA_2) %>%
  filter(Compartment == "Soil") %>%
  select(-Compartment)

## Combine root and soil data for pSEM analysis
microbiome.pSEM.data <- sample.meta.data %>%
  select(Population) %>%
  full_join(root.PCoA.vectors, by = "Population") %>%
  full_join(soil.PCoA.vectors, by = "Population") %>%
  full_join(urbanization.data, by = "Population") %>%
  full_join(carbon.data, by = "Population") %>%
  full_join(nitrogen.data, by = "Population") %>%
  group_by(Population) %>%
  summarise(across(dplyr::everything(), mean), .groups = "keep")
```

```
## Microbiobiome phyloseq
write_rds(
  microbiome.phyloseq.reference,
  file = "data/microbiome_phyloseq_reference.rds"
)

## Microbiome tidyamplicons
write_rds(
  microbiome.tidyamplicon.reference,
  file = "data/microbiome_tidyamplicon_reference.rds"
)
```

```
## Root tidyamplicons
write_rds(
  root.tidyamplicon.reference.rarefied,
  file = "data/root_tidyamplicon_reference.rds"
)

## Soil tidyamplicons
write_rds(
  soil.tidyamplicon.reference.rarefied,
  file = "data/soil_tidyamplicon_reference.rds"
)

## Alpha diversity
write_rds(
  alpha.diversity.cleaned.data,
  file = "data/alpha_diversity_data.rds"
)

## Multivariate dispersion
write_rds(
  multivariate.dispersion.data,
  file = "data/multivariate_dispersion_data.rds"
)

## PCoA (Bray-Curtis)
write_rds(
  BC.distance.PCoA,
  file = "data/BC_PCoA.rds"
)

## PCoA (UniFrac)
write_rds(
  UniFrac.distance.PCoA,
  file = "data/UniFrac_PCoA.rds"
)

## PCoA (weighted UniFrac)
write_rds(
  weighted.UniFrac.distance.PCoA,
  file = "data/weighted_UniFrac_PCoA.rds"
)

## Microbiome pSEM data
write_rds(
  microbiome.pSEM.data,
  file = "data/microbiome_pSEM_data.rds"
)

## GAMMs
save(
  alpha.diversity.cleaned.data, multivariate.dispersion.data,
  BC.distance.PCoA, UniFrac.distance.PCoA, weighted.UniFrac.distance.PCoA,
  ASV.richness.by.distance.GAMM, ASV.richness.by.HII.GAMM, ASV.richness.by.ISC.GAMM,
```

```
  inverse.simpson.by.distance.GAMM, inverse.simpson.by.HII.GAMM, inverse.simpson.by.ISC.GAMM,
  ASV.evenness.by.distance.GAMM, ASV.evenness.by.HII.GAMM, ASV.evenness.by.ISC.GAMM,
  faiths.PD.by.distance.GAMM, faiths.PD.by.HII.GAMM, faiths.PD.by.ISC.GAMM,
  BC.multivariate.dispersion.by.distance.GAM, BC.multivariate.dispersion.by.HII.GAM,
  BC.multivariate.dispersion.by.ISC.GAM,
  UniFrac.multivariate.dispersion.by.distance.GAM, UniFrac.multivariate.dispersion.by.HII.GAM,
  UniFrac.multivariate.dispersion.by.ISC.GAM,
  weighted.UniFrac.multivariate.dispersion.by.distance.GAM, weighted.UniFrac.multivariate.dispersion.by
  weighted.UniFrac.multivariate.dispersion.by.ISC.GAM,
  file = "data_analysis/07-diversity_composition_analyses/diversity_composition-reduced_workspace.RData
)
```

R Session Information

Table 25: Packages required for data management and analysis.

Package	Loaded Version	Date
ape	5.8	2024-04-11
bayestestR	0.13.2	2024-02-12
broom	1.0.6	2024-05-17
correlation	0.8.4	2023-04-06
datawizard	0.10.0	2024-03-26
dplyr	1.1.4	2023-11-17
easystats	0.7.1	2024-03-28
effectsize	0.8.8	2024-05-12
forcats	1.0.0	2023-01-29
ggplot2	3.5.1	2024-04-23
insight	0.19.11	2024-05-12
kableExtra	1.4.0	2024-01-24
knitr	1.46	2024-04-06
lattice	0.22-6	2024-03-20
lubridate	1.9.3	2023-09-27
mgcv	1.9-1	2023-12-21
modelbased	0.8.7	2024-02-15
nlme	3.1-164	2023-11-27
parameters	0.21.7	2024-05-14
performance	0.11.0	2024-03-22
permute	0.9-7	2022-01-27
phyloseq	1.41.1	2024-03-23
picante	1.8.2	2020-06-10
purrr	1.0.2	2023-08-10
readr	2.1.5	2024-01-10
report	0.5.8	2023-12-07
see	0.8.4	2024-04-29
stringr	1.5.1	2023-11-14
tibble	3.2.1	2023-03-20
tidyamplicons	0.2.2	2022-09-10
tidyr	1.3.1	2024-01-24
tidyverse	2.0.0	2023-02-22
vegan	2.6-6.1	2024-05-21